Mining the Sequence Read Archive to identify crAssphage, a ubiquitous inhabitant of the human microbiome, in dog and pig samples

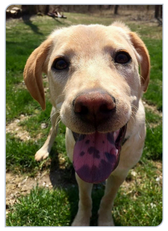
Haley Leffler[1], Sruthi Ganapaneni[1], Bhavya Papudeshi[2], Sheri A. Sanders[2], Thomas G. Doak[2]

[1]Department of Human Biology, Indiana University; [2]National Center for Genome Analysis Support, Pervasive Institute of Technology, Indiana University

Introduction

- There are more bacterial cells (3.8×10^{13}) in the body compared to human cells (3.0×10^{13}).
- Within the gastrointestinal (GI) tract is a gut microbiome, which impacts many factors of health.
  - The gut microbiome includes bacteriophages – viruses that infect bacteria.
- 90% of humans have crAssphage in their gut microbiomes, but the role of the phage is unknown. This ubiquitous phage parasitizes on *Bacteroides intestinalis*, a bacteria that plays a key role in digestion.
- CrAssphage has been found in termite gut metagenomes, humans, and some water samples. CrAss-like genomes have been identified in Old-World and New-World primates, but have not been identified in other species.
- Since the human gut microbiome shares 3.2% of species with dog microbiomes, I hypothesize that crAssphage may be present. Dogs have had close relations to humans for thousands of years and can experience similar diseases humans face, *e.g.* irritable bowel disorder (IBD) and obesity.
- Additionally, 8.2% of the species in the human microbiome overlaps with pig samples, so I hypothesize that crAssphage may be present. The pig GI tract is often used as a model in education because of its close similarity to humans.

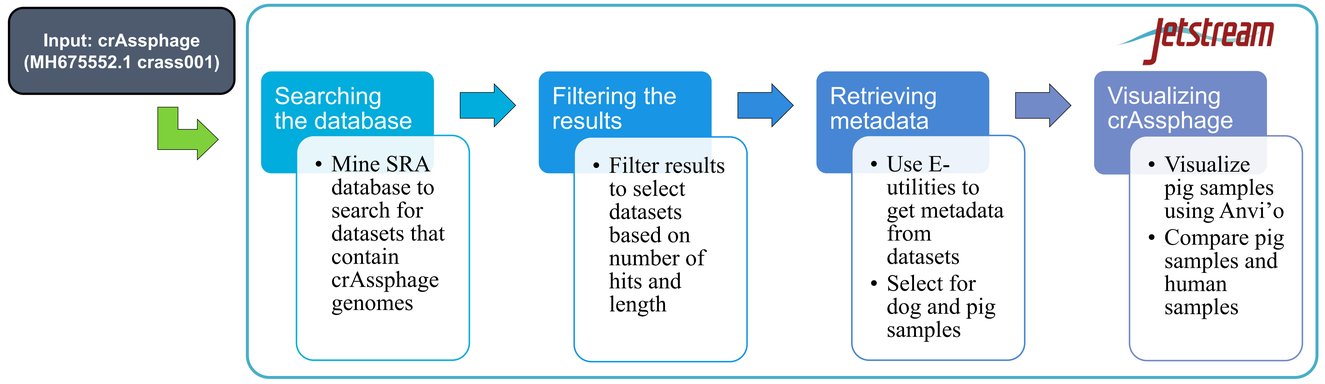
To further explore the presence of crAssphage in dogs and pigs, we developed a workflow using cloud computing infrastructure and publicly available data.



Resources

- The Sequence Read Archive (SRA) is a database that hosts ~14PB of sequencing data with metadata; most published genomics projects around the world have deposited their "raw" sequence data here. SearchSRA is being used for the ability to mine the large database.
- Jetstream is a cloud computing infrastructure that is easy for researchers inexperienced in computing to use because of straightforward user interface.

Workflow



Figures

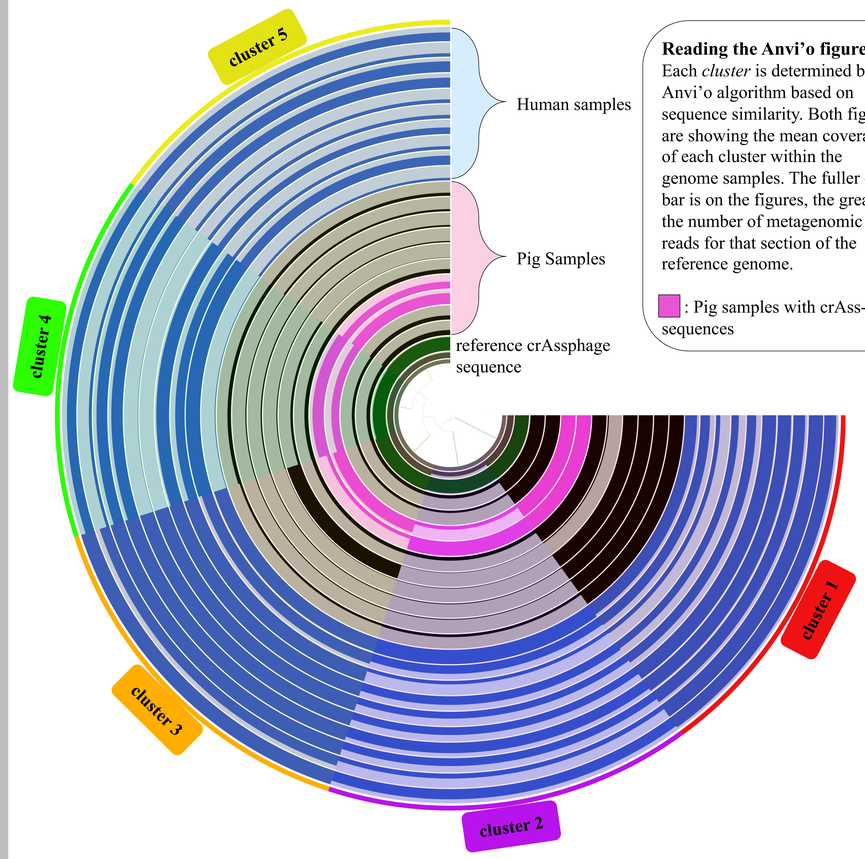


Figure 1: Circular representation showing the genomic distribution of human samples with pig samples with crAssphage.

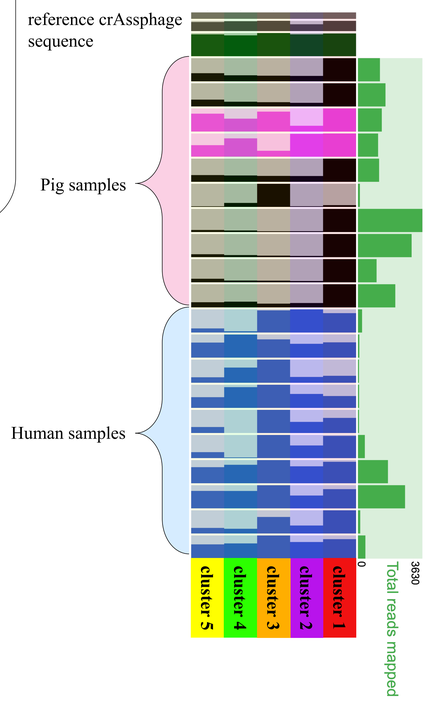


Figure 2: Tree representation showing the genomic distribution of human microbiome samples to pig samples

Results

- We developed a workflow utilizing Jetstream infrastructure to explore the presence of a genome in SRA, and made it available to researchers.
  - Jetstream: https://use.jetstream-cloud.org/application/images/831
  - NCGAS GitHub repository: https://github.com/NCGAS/CEWiT-REU-Identifying-datasets-in-SRA-using-Jetstream
- Our analysis found that **crAssphage is not found in dog samples**, however **crAss-like sequences were found in two pig samples**.
- We found that the majority of the pig samples were false positives, after filtering for alignment length. However, two pig samples were identified to contain crAss-like sequences because of the similarities in coverage compared to human microbiome samples (Figure 2).
- A BLAST analysis was done on cluster 1 for the pig samples to determine the genes within that section. Most of the genes were common genes found in bacteriophages not specific to crAssphage.

Discussion

- The absence of crAss-like sequences from dog samples can be explained by the higher acidity in the stomach and more antibacterial properties of their saliva compared to humans.
- The two pig samples are suggested to contain crAss-like sequences because of the filtering steps that were taken and the similar genomic distribution of crAssphage with the human microbiome.

Using the developed workflow, two pig samples have been found to contain crAss-like sequences which suggests there is a commonality between the pig and human microbiome. This data and workflow is available for public use and can be accessed in GitHub and Jetstream.

<u>Future directions</u>

The two pig samples could be further analyzed to identify the common genes between the pig and human samples to reveal more information on crAss-like sequences that were identified in the two pig samples.

Acknowledgements

Rob Edwards at SDSU